AD-A067 033 AVCO SYSTEMS DIV WILMINGTON MASS F/G 17/1

DEMONSTRATION OF FEASIBILITY OF AVCO DATA ANALYSIS AND PREDICTI--ETC(U)

SEP 71 H E HUNTER, N H KEMP

UNCLASSIFIED AVSD-0411-71-RR-VOL-1 NL

1 OF 1

ADA

067033

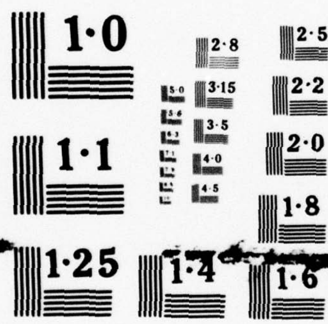
END

DATE

FILMED

6-79

DDC


1·0
2·8
2·5
3·15
2·2
5·0
5·6
3·5
2·0
1·1
4·0
4·5
1·8
1·25
1·4
1·6

LEVEL II -- MOST Project 13

Final Report For:

DEMONSTRATION OF FEASIBILITY OF AVCO DATA ANALYSIS AND PREDICTION TECHNIQUES (ADAPT) FOR SONAR DETECTION

VOL. I: INTRODUCTION, RESULTS, APPENDICES

AVSD-0411-71-RR
September 1971

By

Herbert E. Hunter and Nelson H. Kemp

Prepared For:

JOHN HOPKINS UNIVERSITY
APPLIED PHYSICS LABORATORY
CONTRACT NO. 341787

~~720106-C135~~

Prepared By:

AVCO GOVERNMENT PRODUCTS GROUP
AVCO SYSTEMS DIVISION
WILMINGTON, MASSACHUSETTS 01887

9 Final Report. For:

6

DEMONSTRATION OF FEASIBILITY OF AVCO DATA ANALYSIS AND PREDICTION TECHNIQUES (ADAPT) FOR SONAR DETECTION.

Volume [redacted] I. INTRODUCTION, RESULTS, APPENDICES.

14 AVSD-0411-71-RR-VOL-1

11 Sep[redacted]71

By

10 Herbert E. Hunter [redacted] Nelson H. Kemp

12 42p.

Prepared For:

JOHN HOPKINS UNIVERSITY
APPLIED PHYSICS LABORATORY
CONTRACT NO. 341787

Prepared By:

AVCO GOVERNMENT PRODUCTS GROUP
AVCO SYSTEMS DIVISION
WILMINGTON, MASSACHUSETTS 01887

This report is prepared in two volumes. Volume I is unclassified and contains the Introduction, Conclusions and Recommendations, Description of ADAPT and Appendices A and B. Volume II is classified Confidential, and contains the technical studies. The entire Table of Contents is given in each volume.

ACKNOWLEDGEMENT

J. L. Phillips and R. J. McConahy of the Applied Physics Laboratory have contributed greatly to this work by their interest, encouragement and guidance, as well as by supplying the data upon which the analysis is based. At the Avco Systems Division Mrs. Patricia Gaudet provided invaluable assistance with the computations, and in the preparation of this report. Mrs. Joyce Merritt also assisted with the computations, Dr. L. Meixsell was helpful in discussions on the statistical aspects of the work.

CONTENTS VOLUME I

72^106-^135

CONTENTS VOLUME II

1.0 INTRODUCTION

This report presents the results of a study which showed the feasibility of using the Avco Data Analysis and Prediction Techniques (ADAPT) to improve the ability to detect a sonar target in a sea and self-noise background by more than doubling the detection range compared to the reference conventional detection. The study concentrated on three areas: 1) Comparison of ADAPT methods with time averaging for characterizing the self-noise signals; 2) Comparison of ADAPT methods with time and frequency averaging as a target detector; 3) Determining the effect on detection of using ADAPT's ability to omit the characterizable noise in developing a detection algorithm. These areas will be covered in detail in the subsequent sections of this report.

ADAPT is a method of empirical data analysis developed at Avco Systems Division which is capable of extracting cataloging, sorting, prediction and detection laws out of large volumes of empirical data. A description of the ADAPT methodology, as applied to the present detection problem, is given below in Section 3.0.

The use of an empirical analysis necessitates learning data upon which to base the analysis. For this study, the data was supplied by the Applied Physics Laboratory (APL) of Johns Hopkins University. This data was in the form of a digitized energy density spectrum for each signal, covering the frequency range 0-2 KHz in 512 frequency samples. Thus each frequency bin was 2000/512 = 3.90625 Hz wide. In the time plane, the signals were thus 0.256 sec long. The signals consisted of two classes. One class was actual background and self-noise signals, which will often be referred to as B signals for convenience. The other class was simulated target (T) signals manufactured on a computer by APL personnel. They were made by adding simulated target returns characteristic of a certain submarine to actual B signals, and were produced with several signal-to-noise ratios (SNR). Each T signal was labeled with a value of SNR, to identify how much target was added to the B signal, but this number is not to be taken as the SNR for the whole target spectrum added, as it only refers to the SNR in a small part of the spectrum. It is best looked on as a label which grades (in db) the amount of target added. The target spectrum added was the same for all signals which were used as learning data.

The number of signals of each class and SNR supplied for learning data are shown in the second column of Table 1.1. The development of algorithms was concentrated on the range SNR= -6 to +6, where 200 signals of each were available. The signals at other SNR were used to explore the performance of the resulting algorithms.

A small group of signals, labeled D (for Doppler) in Table 1.1, had a frequency shift in the target spectrum, corresponding to target motion. These were not used as learning data, but were used to test the algorithms to determine the effect of such frequency-shifted signals.

In addition to the data in Table 1.1, a set of unidentified signals were also supplied to Avco in the same format. These are referred to as Category III signals, and were used to test the comparative performance of the ADAPT and conventional averaging methods of detection. These signals were mostly different from the learning signals in that the background signals or target spectra added were varied, rather than just being changed in SNR. Table 8.2 shows that there were 82 groups of Category III signals, of which all but 4 were groups of 10 signals each. The other 4 groups each contained 5 signals, as noted. Table 8.3 groups these into 13 categories of different signals..

The remainder of this report describes the analysis and processing performed by Avco on the data just described. Section 2.0 presents the conclusions and recommendations of this study. After the general description in Section 3.0 of the ADAPT methods of data processing and detection, Section 4.0 presents the methods used for conventional detection by time and frequency averaging, and the results are given for the learning data. Section 5.0 presents the studies of the B signals, and their characterization both by time averaging and by ADAPT. Section 6.0 deals with the development of the ADAPT detection algorithms, and their application to the learning data. In Section 7.0, the ADAPT characterization of the B signals is used to subtract the characterizable noise from the signals so as provide a clearer distinction between the B and T classes; some preliminary results of the use of this technique are given. Section 8.0 shows the comparison of ADAPT and conventional detection on both the learning data and the Category III data.

2.0 CONCLUSIONS AND RECOMMENDATIONS

The studies described in the report show that the ADAPT methods can provide a considerable improvement over conventional energy threshold detection. The degree of improvement increases as the amount of time averaging increases. For no time averaging (i.e. using the energy density spectra of the original 1/4 second signals supplied) ADAPT shows the ability to detect signals 3 to 6 db lower than the best conventional detection method tried, at detection probabilities of 0.1 to 0.2. For high detection probabilities, ADAPT and the best conventional method give about the same results. This is shown in Figure 2.1, where the detection probability P_D is plotted against SNR for ADAPT detection, and for conventional detection over three different frequency bands. The improvement achieved by ADAPT at low SNR is clearly visible.

Time averaging improved all detection methods by reducing random noise. It improves ADAPT detection even more than conventional detection, as shown in Figure 2.2. This figure gives the results of detection algorithms using averages of the energy density spectra of 10 1/4 second signals, i.e., averaging over 2.5 second. Here ADAPT shows an advantage over the best conventional detection of 8 db at detection probability of 0.5, and 6 db at 0.95 and 0.2. Comparison of Figures 2.1 and 2.2 shows that this averaging improves ADAPT detection by 9 to 10 db, while it improves conventional detection by 4 db.

It seems clear from these results that ADAPT can provide more sensitive detection algorithms than the type of conventional energy detection considered here. In characterization of the background signals, ADAPT was able to provide a whole sequence of signals, which can be used to describe the background signals. The first signal in the sequence is closely related to the average signal, and the subsequent signals describe smaller and smaller deviations from the first signal. Any given background signal can be accurately represented as a linear combination of this sequence of signals.

This characterization permitted ADAPT to subtract the "characterizable noise" from both the background and target signals by using the first 3 or 4 terms of the sequence. After this subtraction, the resulting background signals looked very much like random noise, with no noticeable features. The resulting target signals, on the other hand, showed spectra with a definite character, both broad and narrow band, which is associated with the target spectrum used to construct them. The difference between the background and target signals were much more apparent to the eye after ADAPT subtracted the characterizable noise than it was in the original signals.

ADAPT detection algorithms were constructed using these noise-subtracted signals. However, time and cost did not permit complete development of these algorithms, so the results obtained were not as successful as the ADAPT detection on the original signals.

In addition to the resources limitation just mentioned, other limitations prevented ADAPT from showing its full potential. There were only 200 signals at each SNR, and this hampered the use of time averages. For averages of 10, this left only 20 signals, which is too few for ADAPT to use as both learning and proof test data. To cope with this limitation, learning was done on overlapped averages, so successive averaged signals were not independent. Thresholds were set on independent averages, but only 10 were used, and this is too few to yield reliable statistics. In addition, the use of one kind of signal for learning and a second kind for setting the threshold, may have hampered ADAPT's performance. Furthermore, there was no chance to explore further degrees of averaging. There were some 20-average studies, but the use of only 5 averaged signals to set the thresholds clearly made the results unrealistic. The optimal degree of averaging for ADAPT could not be found. Based on the improvement found with 10-averages, it is probably greater than 10. Avco's hypothesis is that ADAPT's extra gain by averaging is related to its ability to use more of the information in the data. If this is correct, the optimal degree of averaging is more than 20.

A third limitation was that no consideration was given to how the ADAPT scheme might best be utilized in actual operation on a submarine. However, the many options of processing available in ADAPT permit algorithm development to be tailored to the operational mode in which it is to be used. In the present study, no advantage could be taken of this flexibility in ADAPT.

The results and limitations of this study lead to recommendations for further studies which will help resolve some of the uncertainties and questions raised by the present study.

The unaveraged algorithm using the noise-subtracted signals should be optimized. This will permit a firm conclusion as to the advantage achieved by noise subtraction on the present data base.

Neither the original nor the noise-subtracted 10-average algorithms were optimized in this study. Further work could be done in this area. However, the learning data set used here is really too small to use for non-overlapped average studies. One way to construct a larger data set for meaningful averaged signals is to combine all the data used in this study, both learning signals and Category III signals. By using all the independent B signals available, and by constructing T signals for each SNR on all these B signals,

almost an order of magnitude larger group of signals would become available. Most of them could be used as learning data, with only a small group of T signals saved for proof test. An alternative way to make a larger data set is to create, either from some other field data on synthetically, a much larger data set. With either of these data sets, the 10-average studies could be carried on more consistently, and extended to a high enough degree of averaging to determine where the advantage of ADAPT over conventional detection steps growing.

Effort should be expended on the relation of operational considerations to ADAPT detection. Such factors as computational requirements, new target response, display of data, differences between patrol and engagement, etc., should be considered.

Another area where ADAPT should be studied is in connection with beam forming. Perhaps ADAPT should use beam-formed data; or perhaps ADAPT can aid in beam forming.

To summarize, ADAPT has demonstrated about an 8 db advantage over conventional energy detection for the problem considered in this study. Some potential for additional improvement with the present data is in the optimization of detection algorithm using noise-subtracted data. There are a number of other areas where ADAPT's usefulness should be explored. In general, ADAPT should be considered an approach for achieving significant improvements in detection range, as well as a tool for analysis of sonar data, both on a submarine and in the laboratory.

3.0 DESCRIPTION OF ADAPT METHODS

The method of detection developed in this report utilizes a data processing method known as Avco's Data Analysis and Prediction Techniques (ADAPT). This method first provides a simple, compact representation of a collection of signals, in which each signal can be represented by a few numbers. Then, this representation can be used in classification schemes for distinguishing between signals from different physical events. These in turn can be fitted into a detection algorithm providing the desired characteristics. Operational application of the detection algorithm requires only a minimum computational capability, since it requires mainly the scalar product of a known algorithm vector with a vector derived from the signal being detected, which requires only a very simple analog circuit or a small portion of a small digital computer. Operational derivation of algorithms on a real time basis can be performed on relatively small computers (i.e., any of the approximately 4K to 8K core size mini computers) when the optimal representation has first been derived using a large computer. This variation in real time onboard data analysis capability with computational requirements, introduces further flexibility into potential use of ADAPT for Sonar data analysis.

This section provides a description of the ADAPT representation, the classification scheme used, and the detection algorithm developed.

3.1 Definition of Data Vectors

The ADAPT techniques address themselves to the efficient representation and classification of data which appears as data vectors, i.e., an indexed series of numbers. In the present case, the data vector is the 512 numbers representing the energy density in the 512 frequency bins from 0-2 KHz, as supplied by APL. Each vector is treated as a vector of N (= 512) dimensions in Euclidean space. In any particular case, if there are M vectors being processed, there is an N x M matrix of numbers. Sometimes, preprocessing is performed on the data vectors before ADAPT is applied, in order to improve the effectiveness of ADAPT. Such preprocessing might include subtracting the average vector (almost always done), normalizing each vector to unit length, taking logarithms, etc. In the present work, the detection algorithms were developed using the log (to the base 10) of the energy density spectrum, which was found more efficient for ADAPT detection.

3.2 Optimal Representation of Data Vectors

With the M input data vectors defined, the first step in ADAPT is to construct from them an orthonormal set of base vectors by the classical Gram-Schmidt procedure. This eliminates any data vectors linearly dependent on others, and results in a set of NC orthonormal N-component vectors, where NC is less than or equal to the smaller of N and M. (The maximum number of linearly independent N-component vectors is N, so if M>N, some of the vectors are surely linearly dependent on others. If M<N, then there are a maximum of M orthogonal base vectors.) The data vectors are now expressed in the Gram-Schmidt base by their components along the NC Gram-Schmidt vectors, so each vector is given by NC components and there are M x NC components altogether. The Gram-Schmidt base vectors themselves have N components. There is usually a reduction in the number of numbers at this stage, since N≥NC, so the M x N original components have been reduced to M x NC. However, there is no reason to believe that the Gram-Schmidt base is the best one for representing the data. It is really an arbitrary orthonormal set of base vectors determined solely by the order in which the data vectors were chosen. The next step is to fine another orthonormal base which is in some sense the best for the given data as a whole.*

To achieve this, a new set of NC N-dimensional orthonormal vectors, rotated from the Gram-Schmidt set, is postulated. This set is to be chosen in an ordered fashion, so that the first vector is the best, and so on. Only a limited number, NR<NC, of these vectors will be used as new base vectors for representing the data vectors. They are chosen as follows: Each data vector is represented by its coefficients in the Gram-Schmidt base, and is projected onto the NR new vectors, giving M x NR components in the new base. If there were as many new vectors as Gram-Schmidt vectors, NR = NC, this would be an exact representation of the data vectors, but since NR<NC, it is only approximate, leaving an error vector as the difference between the data vector and its representation in the new vector base. The square magnitude of this error vector is a measure of the error for each data vector, and the average of these square magnitudes for all data vectors is the mean square error incurred by representing the data vectors in only NR new base vectors.

---

*The approach taken is analogous to the expansion of functions in a set of orthonormal functions, of which Fourier series is the most common example. When one of the classical boundary value problems of mathematical physics is solved, the appropriate differential equation defines a set of orthonormal functions. To satisfy a given function on the boundary, this boundary function is expanded in the set of orthonormal functions obtained. In the present case, there is no differential equation to define a particular set of orthonormal functions. However, it is possible to make this data define its own best set of such functions or vectors.

The new orthonormal set of vectors is chosen by minimizing this mean square error, thus defining the meaning of a "best" set of vectors. If only one vector is used, NR = 1, it is that vector which makes the one-vector representation error the smallest. If a second vector is used also, it is chosen so that together with the first vector, it minimizes the two-vector representation error. This is continued for as many vectors, i.e., as large as value of NR<NC, as is necessary or desirable.

When formulated mathematically, this criterion requires the maximization of a quadratic form whose unknowns are the Gram-Schmidt components of one of the "best" base vectors, and whose coefficient matrix is the covariance matrix of the Gram-Schmidt components of the input data vectors. This problem is a classical one in linear algebra, which often appears under the name of the principal components analysis of a matrix.

The solutions for the unknown best vector components are the normalized eigenvectors of the covariance matrix, and the resulting values of the quadratic form are the eigenvalues of this matrix. Once they are obtained, they are simply arranged in order of decreasing size of the eigenvalues. The largest eigenvalue gives the most reduction in mean square error that can be achieved with only one new base vector; and the corresponding eigenvector in this new base vector. The next largest eigenvalue gives the most reduction in the error that can be achieved by using a second new base vector in addition to the first one found above, and this second vector is the eigenvector of this second largest eigenvalue. This process can be continued until the desired accuracy is achieved. The sum of the NR largest eigenvalues gives the maximum mean square error reduction which can be achieved with NR new base vectors; when adding additional eigenvalues does not significantly increase this sum, the use of the corresponding eigenvectors as additional base vectors does not significantly improve the representation.

A convenient measure of the overall degree of representation achieved with a given number of base vectors is the sum of the eigenvalues of the vectors used, divided by the average square magnitude of the original data vectors. This represents the reduction in mean square error achieved divided by the total error reduction possible; in statistical terms, this is the percent of the variation of the data explained by the representation used.

The degree to which a single data vector is represented in the optimal base is conveniently measured by the ratio of its length along the NR optimal base vectors to its actual length in the original N-dimensional space. If this ratio approaches unity, the vector is well-represented, while if it is

near zero, very little of the vector is included in the optimal space. This also provides a good measure of whether predictions made for a new data vector, using its representation in a given optimal space, are liable to be good. If the vector is as well represented in the space as the data vectors from which the optimal space was constructed, the predictions should be equally as good. Conversely, if it is more poorly represented, the predictions cannot be expected to be as good.

The optimal set of base vectors defined by this procedure is known in the statistical literature as the principal components or Karhunen-Loeve coordinate system. The ADAPT processing of a collection of data vectors yields the components of the data vectors in this optimal base vector system, as well as the components of these base vectors themselves (in the Gram-Schmidt system).

For each data vector, its NR components in the optimal system are the optimal representation of the data in the sense described above. Alternatively, these components may be interpreted as coefficients of the Fourier series of optimal orthonormal functions representing the data vector. This interpretation serves as the basis for the more intuitive description of the ADAPT techniques presented in Appendix A.

The optimal components are used in all further applications of classification and detection. Thus, the original M x N numbers representing M data vectors have been reduced to M x NR components, plus N x NR numbers to define the optimal vector base. Since the base system is optimal, the number of terms, NR, necessary to give a useful representation of a data vector is small, of the order 10, and the reduction in the number of numbers is large, often a factor of 50 to 100.

In the process described so far, the optimal vectors are represented by their NC components in the Gram-Schmidt base, so they are a linear combination of the NC Gram-Schmidt vectors, with their NC components as the coefficients. Since the Gram-Schmidt vectors are N-dimensional vectors in the original space of the data vectors, the optimal vectors can also be represented in this space by performing the linear combination.

3.3 Use of the Optimal Representation for Detection

Having arrived at the optimal (principal component or Karhunen-Loeve) representation, attention is now turned to use of the optimal components for detection. Detection may be looked on as a sorting problem, where each signal is sorted into a background class or a target class. The first step is to derive a sorting algorithm which simply characterizes each signal data vector in a way which provides maximum separation between

the two classes. Then a detection algorithm is developed, based on this characterization, to provide the desired detection scheme.

For sorting, the representation of a data vector as a point in optimal coordinates is used. There are a number of linear sorting schemes which can be applied. The one used here assigned a single number to each vector in the following way: All the vectors are divided into two classes according to the sorting desired. Then an unknown direction (vector) in the optimal space is postulated, and the projection of each vector on that direction is obtained. This projection is a scalar associated with each data vector. The mean of this projection for each of the two classes is found, and then the difference between the two means. Also, the dispersion of the projections of each class about its own mean is found. The postulated direction of projection is determined by maximizing the distance between the mean projections, while holding a linear combination of the dispersions of projections fixed. When the direction of projection is known, the projection of each vector on it is determined. This linear scheme for maximizing the difference between two classes is a generalization of one first suggested by Fisher, and will be referred to as the generalized Fisher linear discriminant.

The projection of each data vector on the Fisher projection vector is the scalar which characterizes the data vector for detection. The remaining task is to set a threshold for this number, to divide background signals from target signals.

For this purpose, note that the projection referred to above is a linear combination of the optimal components of the data vector, with coefficients equal to the components of the separation vector. By retracing the linear transformations which led from the frequency plane to the optimal coordinates, one can express the projection as a linear combination of components of the original data vector, whose coefficients are the components of the frequency plane expression for the separation vector. This latter vector is usually called the relative importance vector. Thus, the projection may be looked on as a linear combination of the energies at each frequency (or their log). This takes the place of the frequency-averaged energy over a band used in conventional energy detection, which is also a linear combination of energies. Therefore, the same detection theory may be applied to the ADAPT projection as to the average energy in broadband detection. One can appeal to the Central Limit Theorem in both cases, to conclude that both the average energy and the ADAPT projection are normally distributed for large samples. Therefore, the threshold for ADAPT detection has been set using the mean and standard deviation of the Fisher projection in the same way as described for conventional broadband detection using the mean and standard deviation of the broadband energy

in Subsection 4.1 below. Equation (4.3) is used, and the same values of β as given in Table 4.1. The thresholds themselves will be different from those of conventional broadband detection, because the means and standard deviations of the Fisher projection will be different from those of the average energy.

Once the projection vector and threshold are formed, it is not necessary to actually find the optimal coefficients of a new data vector which is being investigated. The transformation from the N-dimensional data vector space to the NR-dimensional optimal vector space can be inverted and incorporated into the detection algorithm. Then the process of applying this algorithm to a new data vector involves primarily the scalar product with this N-dimensional algorithm vector, a rather simple procedure.

A graphical outline of the ADAPT procedures, from representation through detection, is presented in Figure 3.1.

3.4 Advantages of ADAPT

There are many advantages to using the ADAPT data representation before proceeding with any empirical analysis. The most obvious advantage is that the amount of data which must be processed through the detection analysis scheme has been significantly reduced, resulting in both a reduction in the computer time required and the added ability to handle significantly larger quantities of data for a fixed computer size and accuracy.

A more subtle advantage is that this reduction in amount of data may be viewed geometrically as a reduction of the dimensionality of the space in which the detection analysis is carried out. This enables one to handle problems with a smaller number of learning data vectors than it would be possible in the original space, and still avoid what might be termed over-determination of the problem. If the number of learning data vectors is not considerably larger than the dimensionality of the space in which the analyses is performed, there is a very high probability that the law generated will be the peculiar to particular learning data set used, and not generalizable to other data sets.

Another advantage of the ADAPT approach is that each dimension in a detection problem contains the most significant information from all of the components of the initial data vector. In more classical approaches each component must be considered sequentially and independent of the other components. The ADAPT compressed representation of the data insures that the least correlated data is eliminated prior to the development of any algorithms. (See Appendix B.) This results in early elimination of whatever noise is present in the data. Although it is conceivable that useful

information is contained in the poorly correlated portion of the data, the ability of any technique to extract this information decreases rapidly as the data becomes less correlated. The degree of correlation to be required for any given problem is an input to the ADAPT processing.

The procedure of first finding the optimal representation also offers the advantage of an independent validity criteria for the empirical data analysis.

The ADAPT approach further offers significant advantages to the operational implementation of real time sonar signature analysis. The most obvious of these advantages is the uniform and simple format of all of the ADAPT algorithms. This means that the algorithm can be implemented with very small computers or simple analog circuits. Furthermore, ADAPT can provide the operator with the "best" possible two, three or NR-dimensional display of all of the available data. Finally, the small dimensionality required to perform empirical pattern recognition and regression analysis, using the optimally represented data, allows one to supplement these "best" data displays with a real time capability to derive classification laws or to determine the significant characteristics of any new signal which is observed. This can be accomplished with a relatively small computational capability such as that provided by a typical "mini-computer" of approximately 4K to 8K core size.

In summary, the ADAPT approach leads to a very efficient empirical analysis based on an optimal representation of the data which: 1) significantly reduces the number of numbers required to represent any given set of information; 2) eliminates redundant data; and 3) significantly reduces the amount of uncorrelated data. When combined with the detection techniques incorporated in the ADAPT system of programs, there is a significant reduction in manpower, computer hours and roundoff errors with a simultaneous increase in the probability of finding a simple, meaningful empirical relationship.

TABLE 1.1

SIGNALS SUPPLIED FOR LEARNING DATA

SNR (DB)	No. of Signals	No. of 10- Average	No. of 20-Average
B	200	20	10
-21	40	4	2
-15	40	4	2
- 9	40	4	2
- 6	200	20	10
- 3	200	20	10
0	200	20	10
3	200	20	10
6	200	20	10
9	40	4	2
D-3	10	1	-
D 0	10	1	-
D 3	10	1	-
D 6	10	1	-
D 9	10	1	-

FIG 2.1 COMPARISON OF DETECTION PROBABILITY FOR CONVENTIONAL AND ADAPT 1-AVERAGE ALGORITHMS

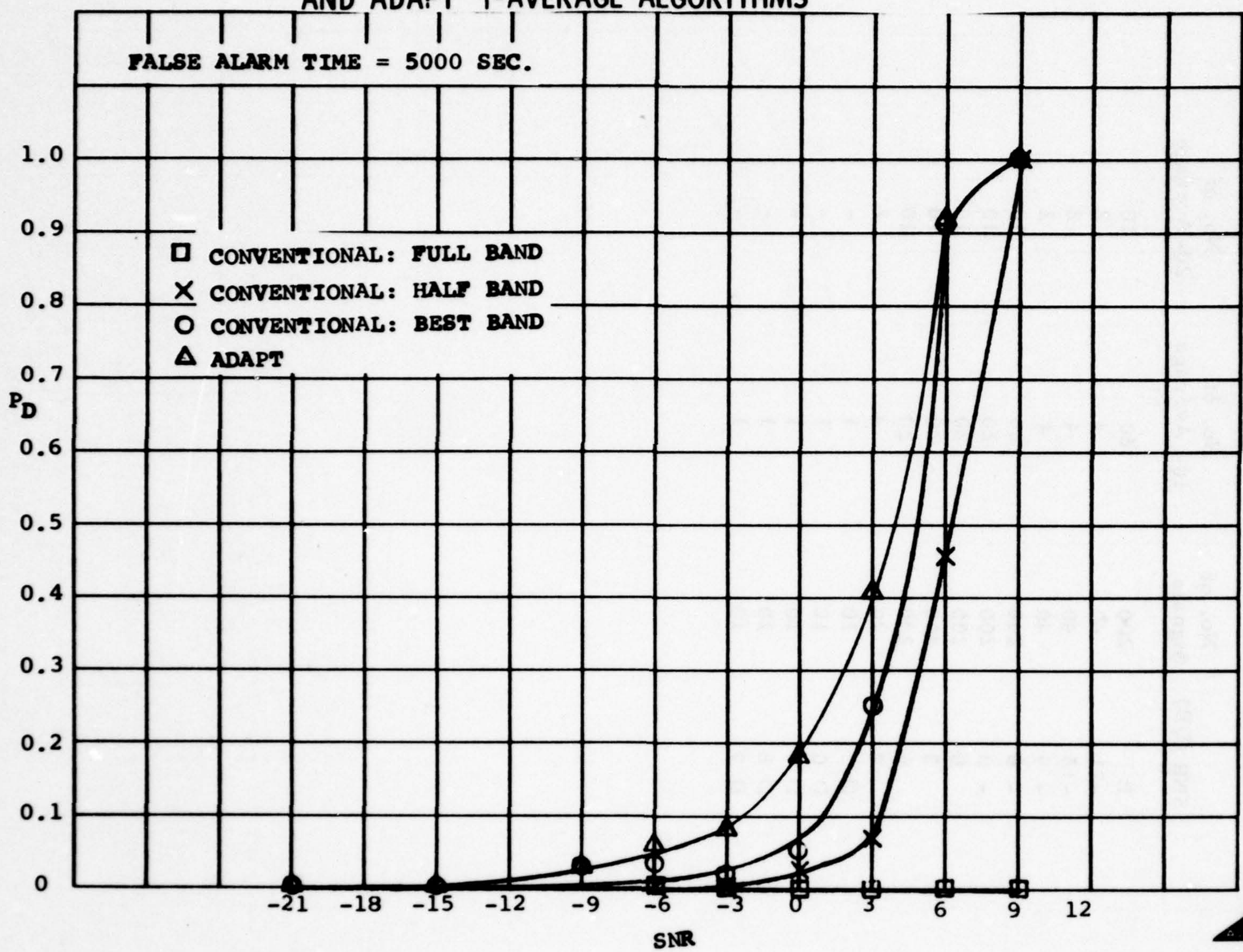
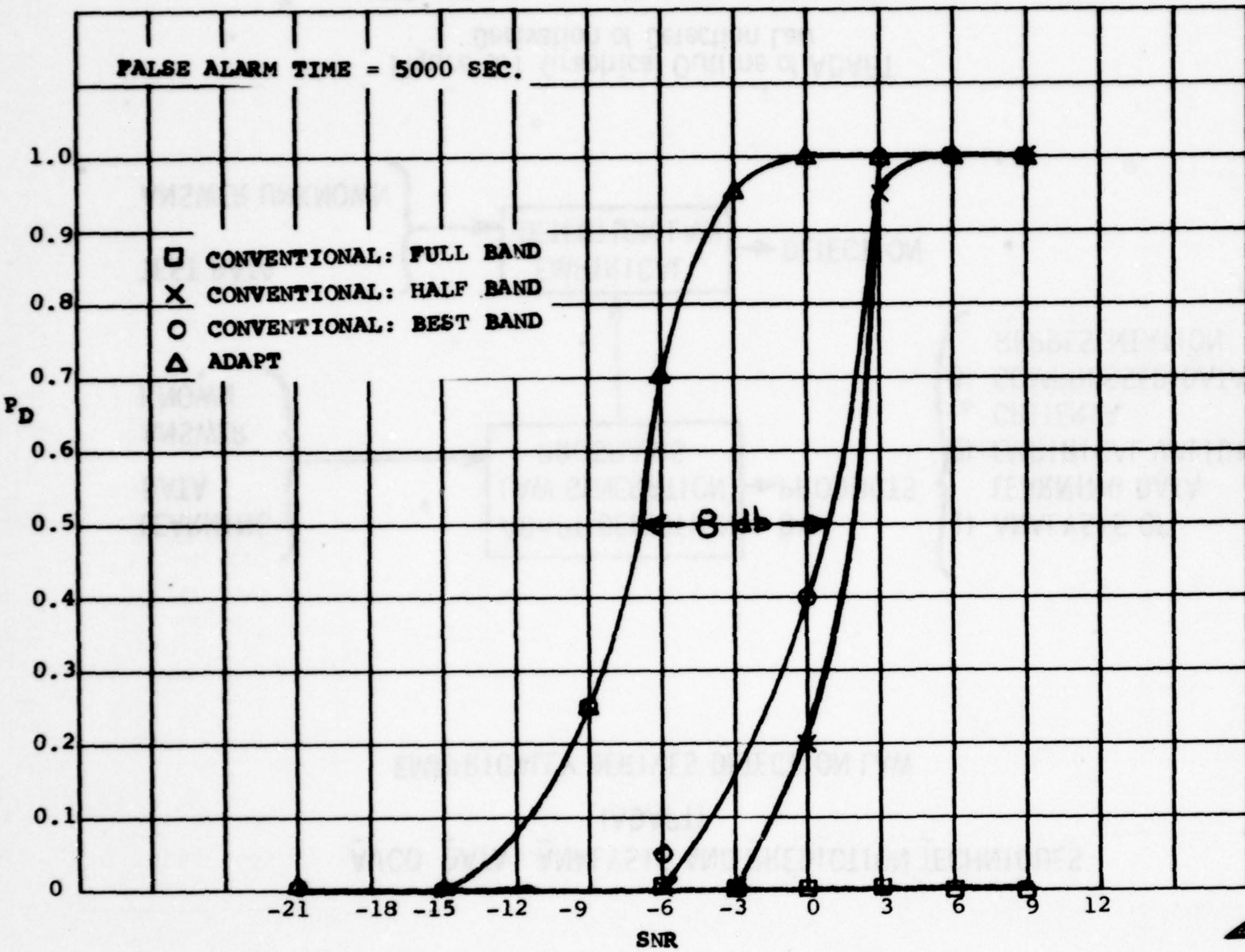


FIG 2.2 COMPARISON OF DETECTION PROBABILITY FOR CONVENTIONAL AND ADAPT 10-AVERAGE ALGORITHMS



AVCO DATA ANALYSIS AND PREDICTION TECHNIQUES
(ADAPT)

EMPIRICALLY DERIVES DETECTION LAW

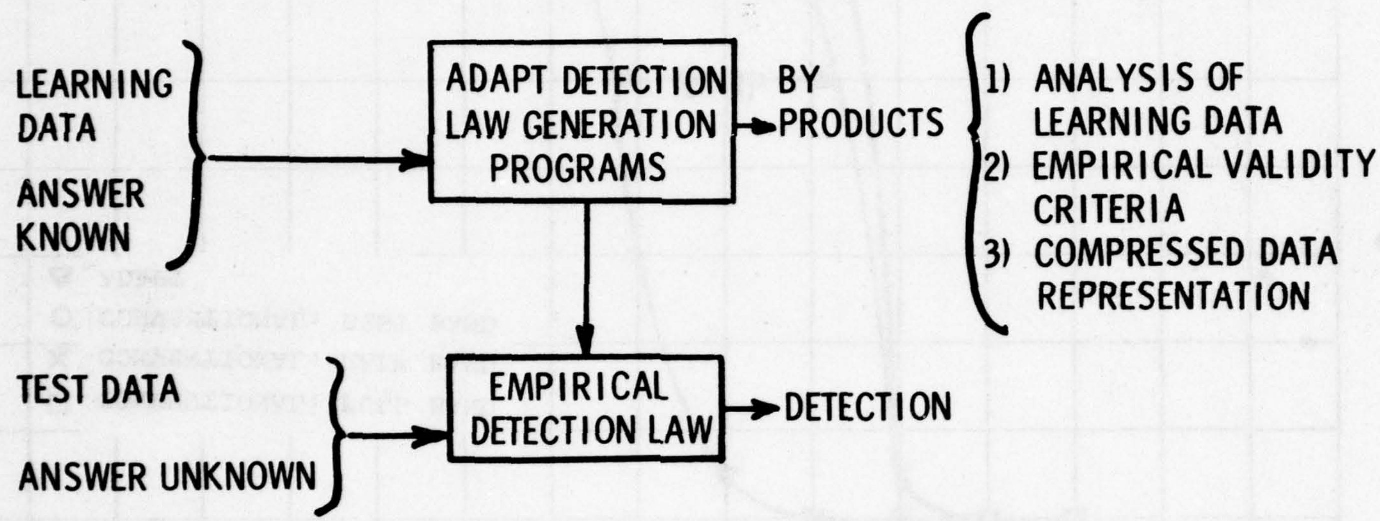


Figure 3.1 Graphical Outline of ADAPT Derivation of Detection Law

1-2555
6-69

APPENDIX A

DESCRIPTION OF ADAPT

INTRODUCTION

Avco's Data Analysis and Prediction Techniques (ADAPT) have been developed over the past 5 years and provide Avco with a unique capability for the application of pattern recognition techniques to empirical data analysis and predictions involving large quantities of data. These techniques were developed as a design tool to evaluate the effectiveness of various decoy concepts for simulating ICBM warheads but are applicable to a large variety of problems. The missile discrimination problem is characterized by the requirement to process very large amounts of data in extremely short periods. This requirement led to the development of a series of computer programs which process a very large quantity of data by compacting it into both an economical and a mathematically more convenient format. This reduced data is then further processed in one of the following processing modes: 1) cataloging, 2) sorting, and 3) parameter prediction.

These techniques are now available as a flexible series of programs. In general, one portion of this series of programs is utilized to generate simple ways of implementing various pattern recognition and prediction schemes based on a known set of learning data. Although a large computer is required to generate these results, the programs that are actually employed for sorting, cataloging and parameter prediction are relatively simple and are compatible with many field-sized computers. These techniques are particularly useful for those applications which require either: the reduction of extremely large quantities of data, a rather simple formulation of algorithms for use in simpler field computers, or for which complex interrelationships are to be determined. By properly formulating the problem these existing programs developed,as part of Avco's missile system technology, are directly applicable to a great many different problem areas.

The following will briefly describe the ADAPT approach and the procedure for formulating empirical problems in a form suitable for analysis with existing ADAPT programs.

1-2555
6-69

ADAPT APPROACH TO EMPIRICAL DATA ANALYSIS

The ADAPT approach to performing empirical data analysis may be divided into four parts: 1) data conditioning, 2) cataloging, 3) sorting, and 4) predicting. The first step in any ADAPT solution is to properly condition the data. This may then be followed by any one of the three remaining tasks listed above. Figure 1 illustrates the fundamental ADAPT assumption concerning data. ADAPT assumes that the data which it is given characterizes some physical phenomena. Here we see that the first observable history, O_1, represents some physical phenomena, say the acoustic signature of a ship moving in calm water. The second observable history, O_2, might be the acoustic signature of the same ship moving in rougher water. A set of many such histories represents the acoustic signature of the ship under varying conditions. A second class within this data might be a series of acoustic signatures representing a whale under various different conditions. An example of this case might be illustrated by the last history of Figure 1, O_m.

Data Conditioning

The data conditioning performed by the ADAPT programs is based on the idea that the standard processing of data often rests on the orthogonality properties of the trignometric functions. That is, data is often represented by trignometric functions which allow certain special processing such as Fourier transforms. These useful properties of trignometric functions exist for any member of the infinite set of orthogonal functions. Examples of the better known are: Bessel Functions, Legendre Polynomials, Jacobi Polynomials, Chebyshev Polynomials, and Hermite Polynomials. In classical boundary value problems the governing differential equation is known and the particular set of orthogonal functions to be utilized in analyzing a given problem is derived from the form of the governing differential equation. However, in the case of a general set of data obtained by making measurements on a phenomena for which the governing equations are not entirely understood, the correct set of orthogonal functions to use is not known. The basis of the ADAPT data conditioning lies in determining the answer to the question: <u>for the given set of data which is to be analyzed what set of orthogonal functions best represents this particular data</u>?

Thus the ADAPT data conditioning procedure begins with an examination of the data to be processed to determine the optimum set of orthogonal functions for representing this particular data set. Although these functions cannot be specified analytically, there are classical numerical techniques available for determining them for any given set of data. In the ADAPT programs this is accomplished by first applying the classical Gram-Schmidt[1] orthogonalization procedure to the data to be analyzed, and following this by Karhunen-Loève [2] type principal component analysis.

This procedure may be looked upon as follows: the Gram-Schmidt procedure is essentially a method of arbitrarily forming a set of orthogonal functions to represent the data. This Gram-Schmidt set of orthogonal functions is arbitrary in that it is entirely dependent upon the order in which the data histories are taken. In essence the procedure consists of taking the first history as the first of the orthogonal functions to be determined and then considering the first history and the second history together to determine two orthogonal functions which represent these two histories. These two orthogonal functions are then considered in conjunction with the third history in the data set and a third orthogonal function selected such that all three of the observable histories are now represented by these three orthogonal functions. This procedure is continued until all of the observable histories have been examined and a set of orthogonal functions to represent these histories has been determined.

If all of the histories are linearly independent the Gram-Schmidt procedure will find a new orthogonal function for each of the histories; however, any history which is linearly dependent on the ones already used will not require an additional orthogonal function, and the number of orthogonal functions found from the Gram-Schmidt procedure will be less than the number of observable histories in the original set.

---

(1) Nering, Evar. D., "Linear Algebra and Matrix Theory," John Wiley & Sons, 1963, pp. 148-149.

(2) Courant and Hilbert, "Methods of Mathematical Physics-I," Interscience Publishers, N. Y., 1963, pp. 23-27.

The optimum orthogonal expansion is now found by determining a new set of orthogonal functions such that the first orthogonal function is selected to contain the most information possible concerning the entire ensemble of observable histories. Then the second orthogonal function is selected to contain the next most information and this process is continued until the last orthogonal function, which contains the least amount of information which was in the original data set, has been determined.

The completion of the Gram-Schmidt and the optimization procedures results in a numerical definition of the optimum orthogonal functions for representing the data set to be analyzed. This data set is then expanded in terms of these orthogonal functions so that each history in the original data set may now be represented by the n coefficients in the series which represents that history. Since these orthogonal functions are optimum for the particular data to be considered, any truncation of this series will result in the best possible representation of the original histories for the number of terms retained.

The preceding discussion has pointed out that there are at least three significant advantages to this new representation of the data which will be summarized here: 1) There is a maximum amount of information contained in any specified number of numbers selected to represent the original data set. This almost always results in a significant reduction in the number of numbers required to accomplish a given analysis task 2) The data to be analyzed throughout the remainder of the ADAPT programs is in good form (i.e., the data contains no singularities) since all of the linear dependence has been removed. Often the remaining process involves orthogonal matrices which have the added advantage that the transpose is equal to the inverse. 3) Examination of the first orthogonal function for any given set of data results in a characterization of the data which may be used to illustrate the typical characteristics of the data set under examination.

Figure 2 summarizes this procedure and outlines the analysis which can be carried out following this data conditioning for the particular set of data which was illustrated for Figure 1. In actuality this set of data consisted of twenty-nine data histories, all similar to the three illustrated

in these figures. The center block of Figure 2 illustrates the first, or most important, of the optimum orthogonal functions used to represent this data. Thus the data conditioning is summarized by the set of optimum orthogonal functions and the coefficients for expanding each history in the data set in a series of these functions. A very useful additional ADAPT output is the average amount of information retained when a given number of terms is used.

Analysis

Companion programs are now operational at Avco to carry this conditioned data through three different types of analysis procedure, namely, sorting, prediction, and cataloging. An important and unique output of ADAPT for sorting and prediction is the relative importance of each of the indexing parameters to the algorithm obtained. This allows one to relate the results obtained back to the physics of the problem by indicating what regions of the independent variable contain the most information for the particular prediction or sorting operation which is to be carried out. In the following paragraphs we shall discuss each of these analysis techniques separately.

A. Cataloging

Figure 3 is an illustration of cataloging and similarity assessment. In the simplest case we may consider each history as defined by its first two coefficients. In this case we may create what is known as a scatter plot which is simply the representation of each history as a single point in the coordinate system which is made up of the first and second coefficients. Thus, each of the three histories previously shown are represented as a single point in this plot.

If a large percentage of the information is contained in the first two terms then the nearness of the points of this plot is the measure of the similarity of the two histories. However, one may proceed one step further in that the scatter plot may be generalized to any number of dimensions. The Euclidian distance between each of these points still represents the similarity of the two histories. In fact it can be shown that if normalized coefficients are used the Euclidian distance between any two points is simply related to the correlation of the two histories represented.

One may represent each history by its coordinates in coefficient space, and when a new history is observed it may be placed in this same coefficient space and its distance from all of the other points calculated. The point to which it is closest represents the history which is most like the new history, and the catalog look-up is completed.

B. Sorting

This same idea of representing each history as a point in coefficient space lends itself to the development of sorting techniques. Figure 4 illustrates one of the linear classification schemes which are currently available in the ADAPT program. This is a scheme which seeks to find a single number which best represents each history with respect to a particular classification which is desired. This is accomplished by projecting the learning data on a direction in coefficient space, which is selected such that the distance between the means of the classes is maximized while the intraclass dispersion is held fixed. When this direction has been selected both the learning data and any new test cases are projected on this direction. If the new test case falls within the limits established by either class defined by the learning data, the test data is assigned to that class. The results of this procedure are primarily to reduce each history to a single number, as illustrated in Figure 5, which empirically best characterizes that history for the particular classification desired.

C. Prediction

The third set of programs available within the ADAPT scheme are for parameter estimation or prediction. Figure 6 illustrates the empirical parameter estimation concepts employed in ADAPT. Again there exists a set of learning data consisting of observable histories. The learning data is first processed through the data conditioning, yielding coefficients and optimum orthogonal functions. The coefficients are then combined with the parameters of the learning data to determine an empirical model. This model relates the parameters to be estimated for each history to the coefficients of that history by least squares analysis. A new test observable history may now be considered. The optimum orthogonal functions obtained from the learning data are utilized to obtain the expansion coefficients for the new data. When these coefficients are inserted into the empirical model it yields the value of the parameter for the test data.

1-2555
6-69

Figure 7 summarizes the entire ADAPT scheme of empirical signature analysis. The ADAPT analysis begins with a set of learning data which is first processed through the ADAPT data conditioning programs. This yields a conditioned data output consisting of coefficients and optimal orthogonal functions. These coefficients are then processed through the various ADAPT algorithm-generation programs which produce cataloging, sorting, and prediction algorithms. These algorithm generation programs also provide plots of the relative importance of each value of the indexing variable in the original observable history to the particular algorithm derived. The algorithms which are generated use only simple mathematical operations and therefore are easily implemented under field conditions. The final step is to process an unknown data sample through whichever algorithm is appropriate for the task required.

FORMULATION OF ADAPT PROBLEM

The key to formulating a problem for the ADAPT process is to understand what is meant by an observable history. Briefly an observable history may be defined as an indexed sequence of numbers which characterize a physical phenomena. This definition covers the normally acceptable observable histories, such as velocity as a function of altitude for reentry vehicles, or vibration amplitude as a function of time for many vibration diagnostic problems, or voltage as a function of time for instruments which measures some time dependent function.

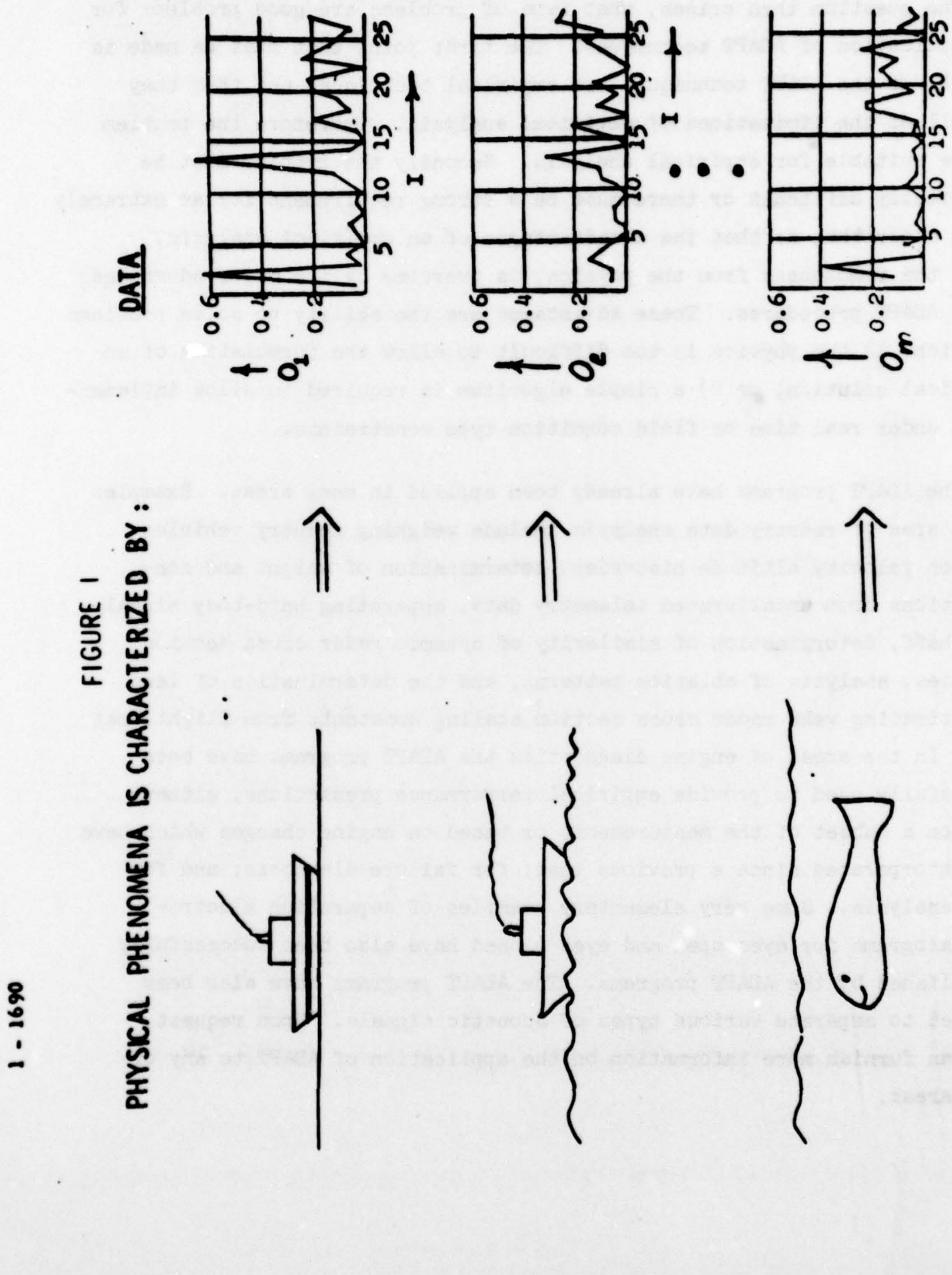
An observable history may also be very different from time like histories. That is, the observable histories may be made up of measurements representing the same physical phenomena. An example of this is illustrated in Figure 8. In this figure we see that an observable history has been constructed from 28 discrete measurements. These measurements taken on a gas turbine engine were characteristic of this engine under a given set of conditions. Thus the first term in the observable history, instead of being the value of a voltage, vibration, velocity, etc., at time 1, was the value of the compressor discharge temperature; the value of the second term in the observable history was the value of the combustor static pressure. Clearly each of these points in this observable history could have been a time history; for example the time history of the compressor discharge temperature could be followed by a time history of the combustor static pressure, which could be followed by a time history of vibration displacement for the power turbine, until a single observable history is constructed from all 28 time histories. This would lead to a long observable history, since it would still be an indexed sequence of numbers characterizing a particular gas turbine engine under a particular set of conditions. The ADAPT programs can handle any number of such data histories consisting of up to 2,000 index points each. By repeated application even the restriction of 2,000 index points can be removed.

In summary the ADAPT procedures are capable of handling in their present form any set of measurements which characterize a physical phenomena. One must then conclude that the ADAPT procedures are capable of processing data for almost any problem in which the physical characteristics of interest

to the analysis influence an indexed sequence of numbers.

The question then arises, what type of problems are good problems for the application of ADAPT techniques? The first point that must be made is that all of the ADAPT techniques are empirical techniques and thus they have all of the limitations of empirical analysis. Therefore the problem must be suitable for empirical analysis. Secondly the problem must be sufficiently difficult or there must be a strong requirement for an extremely simple algorithm, so that the disadvantages of an empirical analysis, namely the remoteness from the physics, is overcome by the major advantages of the ADAPT procedures. These advantages are the ability to solve problems for which: 1) the physics is too difficult to allow the formulation of an analytical solution, or 2) a simple algorithm is required to allow implementation under real time or field condition type constraints.

The ADAPT programs have already been applied in many areas. Examples in the area of reentry data analysis include weighing reentry vehicles based on velocity altitude histories, determination of weight and configurations from uncalibrated telemetry data, separating hard-body signal from chaff, determination of similarity of dynamic radar cross section histories, analysis of ablation patterns, and the determination of laws for estimating wake radar cross section scaling constants from flight test data. In the areas of engine diagnostics the ADAPT programs have been successfully used to provide empirical performance predictions, either based on a subset of the measurements or based on engine changes which have been incorporated since a previous test; for failure diagnosis; and for trend analysis. Some very elementary examples of separating electroencephalograms for eyes open and eyes closed have also been successfully accomplished by the ADAPT programs. The ADAPT programs have also been utilized to separate various types of acoustic signals. Upon request, Avco can furnish more information on the application of ADAPT to any of these areas.



1 - 1690
FIGURE 1
PHYSICAL PHENOMENA IS CHARACTERIZED BY:
DATA
0.6
0.4
0.2
O_1
O_2
O_m
5
10
15
20
25
I

1-1668

FIGURE 2

ADAPT SUMMARY

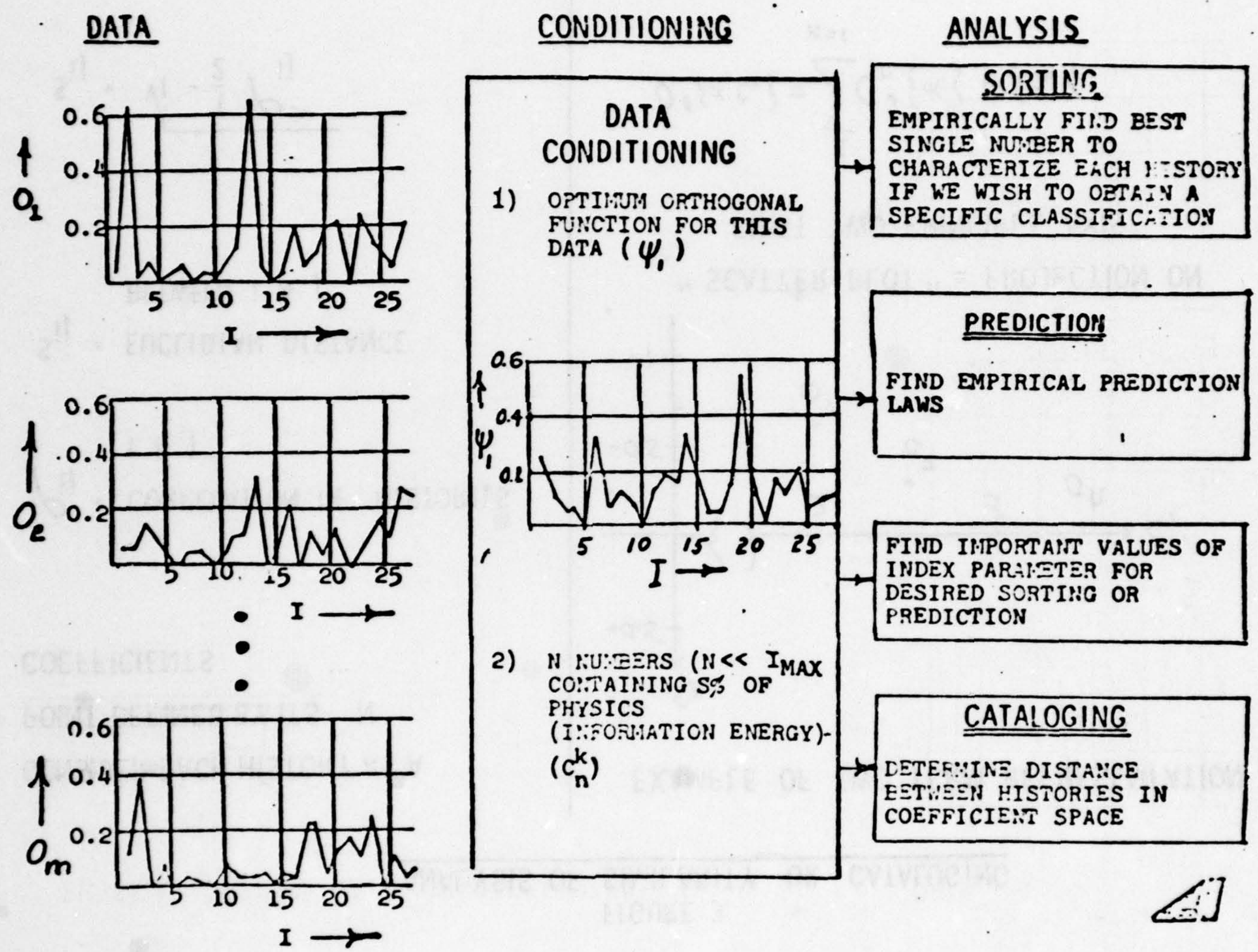


FIGURE 3

ANALYSIS OF SIMILARITY OR CATALOGING

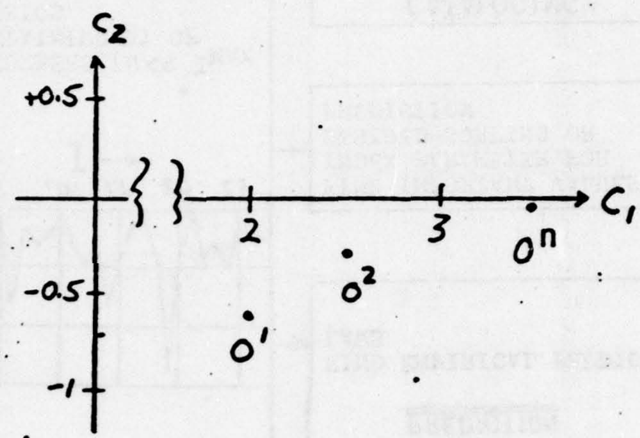
CONSIDER EACH HISTORY AS A POINT DEFINED BY ITS N COEFFICIENTS

ρ_{ij} = CORRELATION OF HISTORIES i & j

S_{ij} = EUCLIDIAN DISTANCE BETWEEN i & j

$$S_{ij} = \sqrt{1 - \frac{1}{2}\rho_{ij}^{\alpha}}$$

EXAMPLE OF TWO TERM REPRESENTATION



" SCATTER PLOT " = PROJECTION ON FIRST TWO PRINCIPLE AXIS

$$O^j\{\alpha, t_k\} = \sum_{n=1}^{2} C_n^j\{\alpha\}\,\Psi\{t_k\}$$

FIGURE 4

PATTERN RECOGNITION

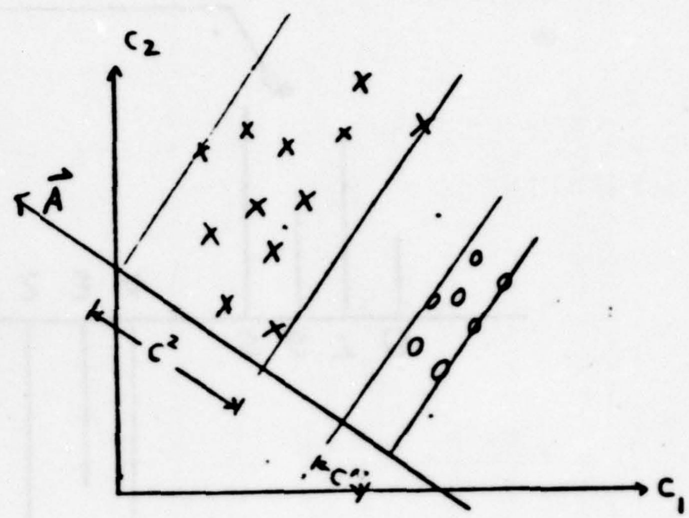
LINEAR CLASSIFICATIONS

- PROJECT ALL POINTS ON DIRECTION $\vec{A}$.
- . IDENTIFY CLASSES BASED ON PROJECTION.
- . EXISTING OPTIONS:

H - OP: MAXIMIZE DISTANCE BETWEEN $\overline{X}_1$ & $\overline{X}_2$.

FIX $\sigma_1 + \sigma_2$

H - OP IN 2 - DIMENSIONS



$\overline{X}_i$ MEAN i th CLASS.

σ_i DISPERSION OF i th CLASS.

FIGURE 5
SORTING RESULTS

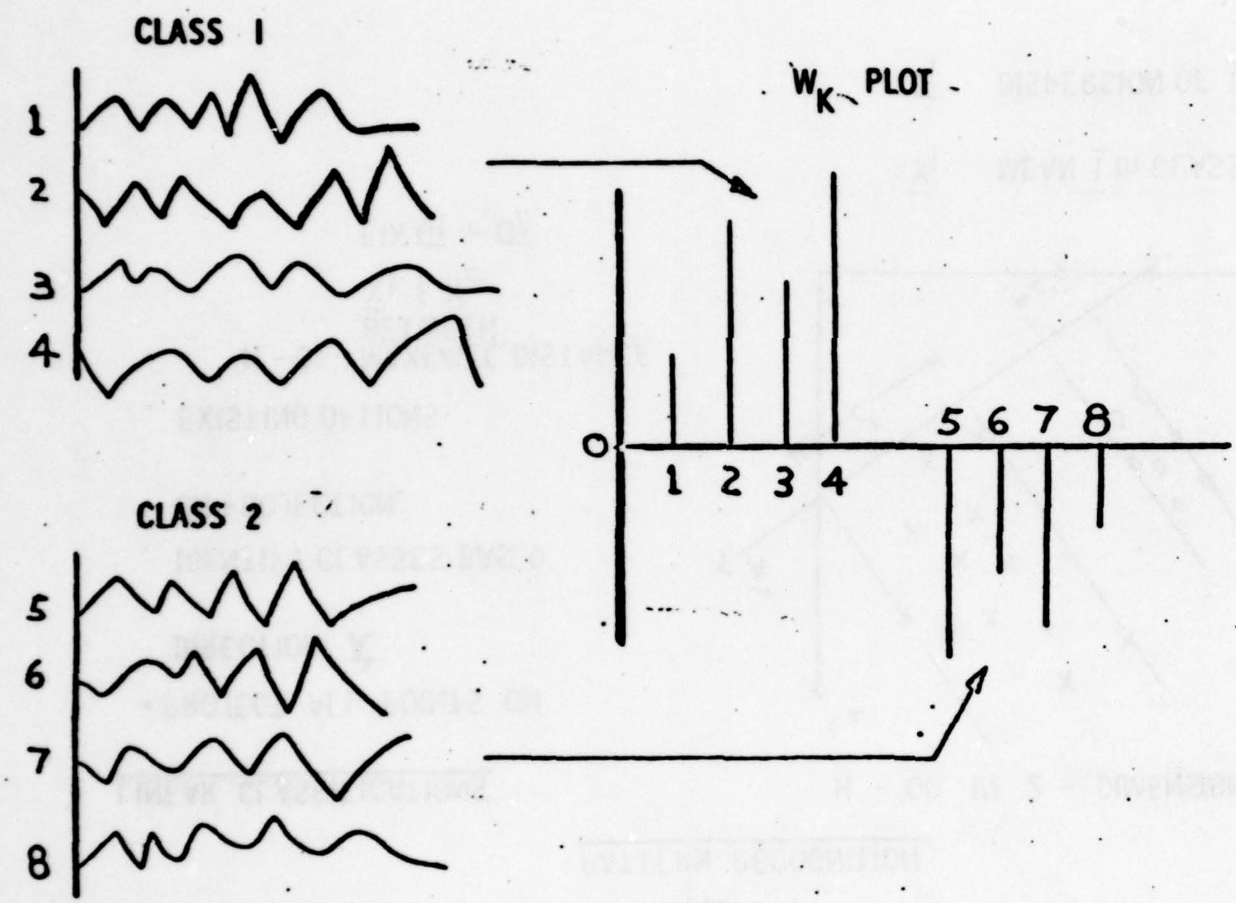
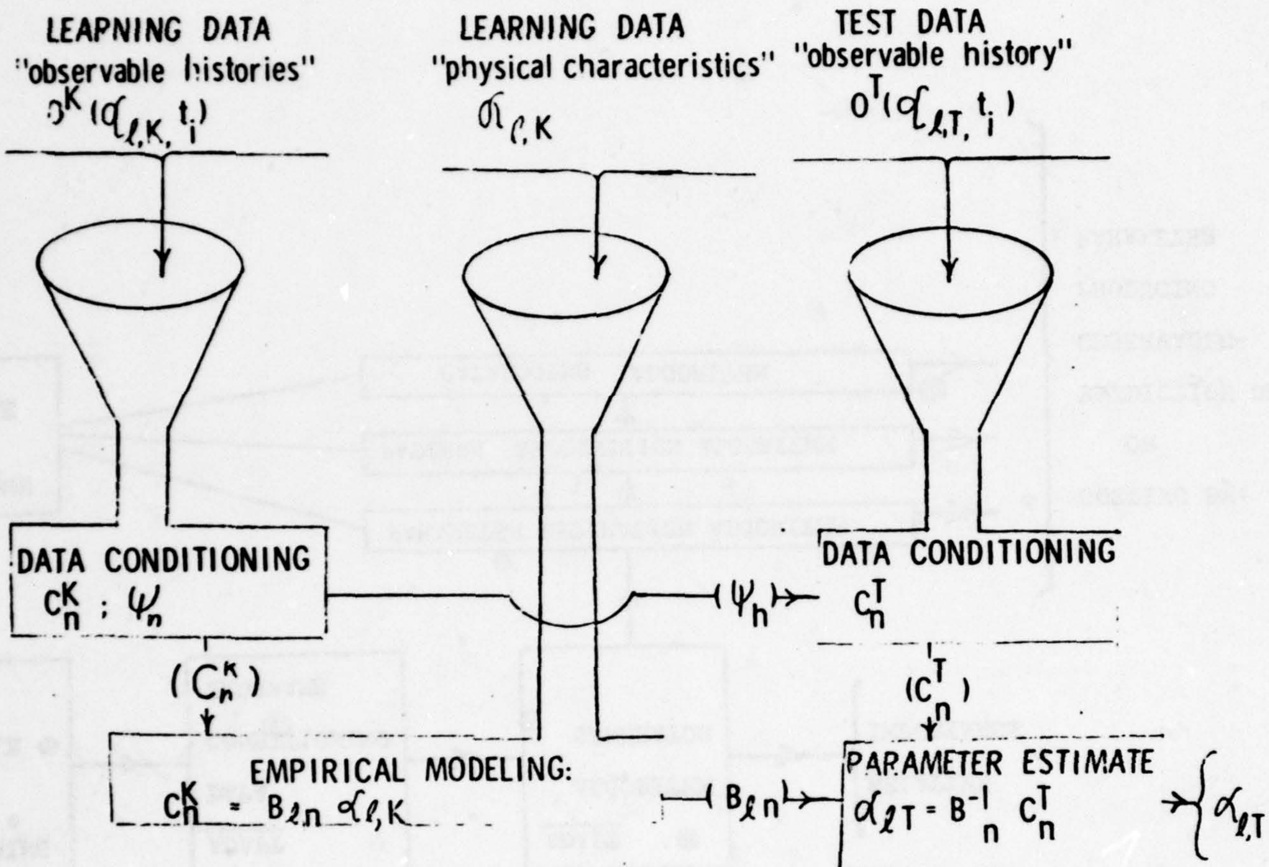


FIGURE 6

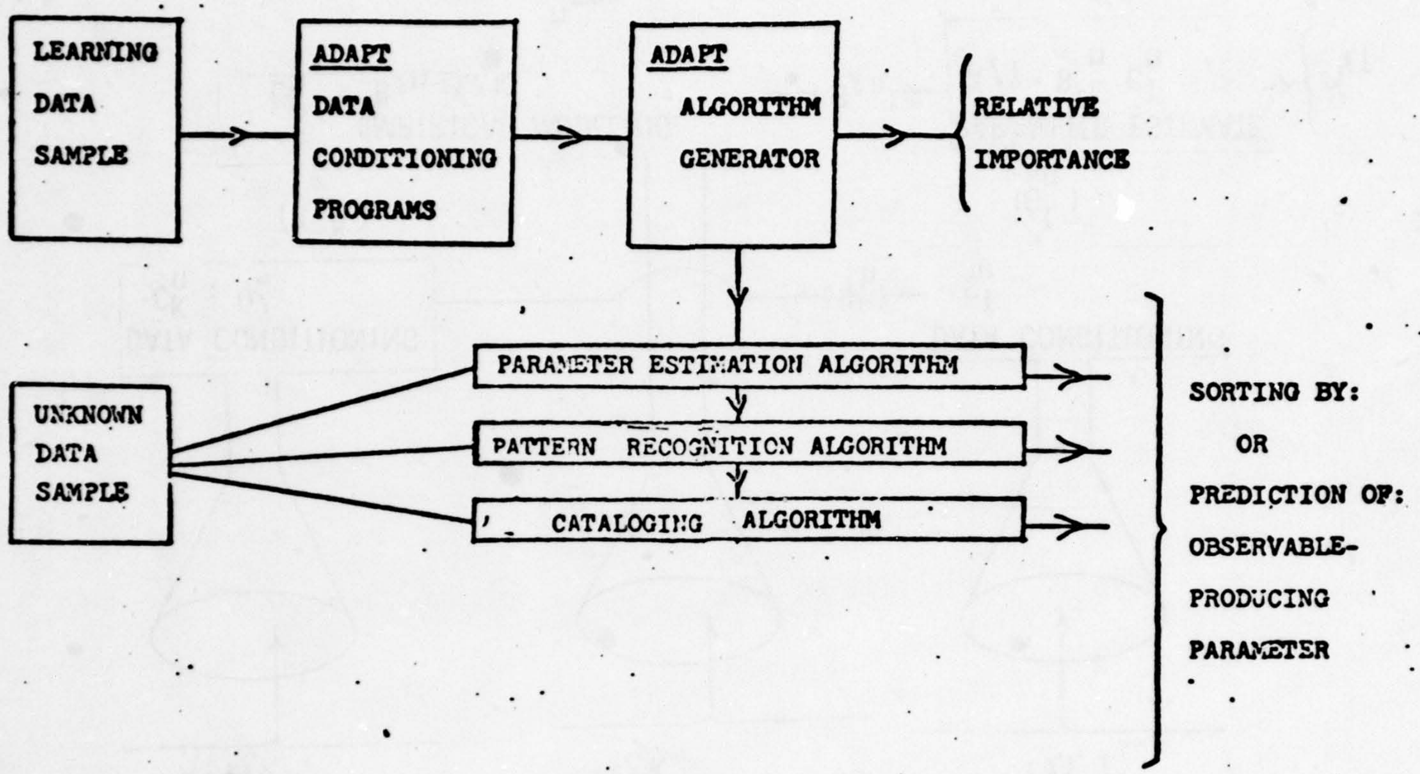
PARAMETER ESTIMATION CONCEPT



$$O^K\{\alpha_{\ell,K}, t_i\} = \sum_{n=0}^{n_{max}} C_n^K\{\alpha_{\ell,K}\}\,\psi_n\{t_i\}$$

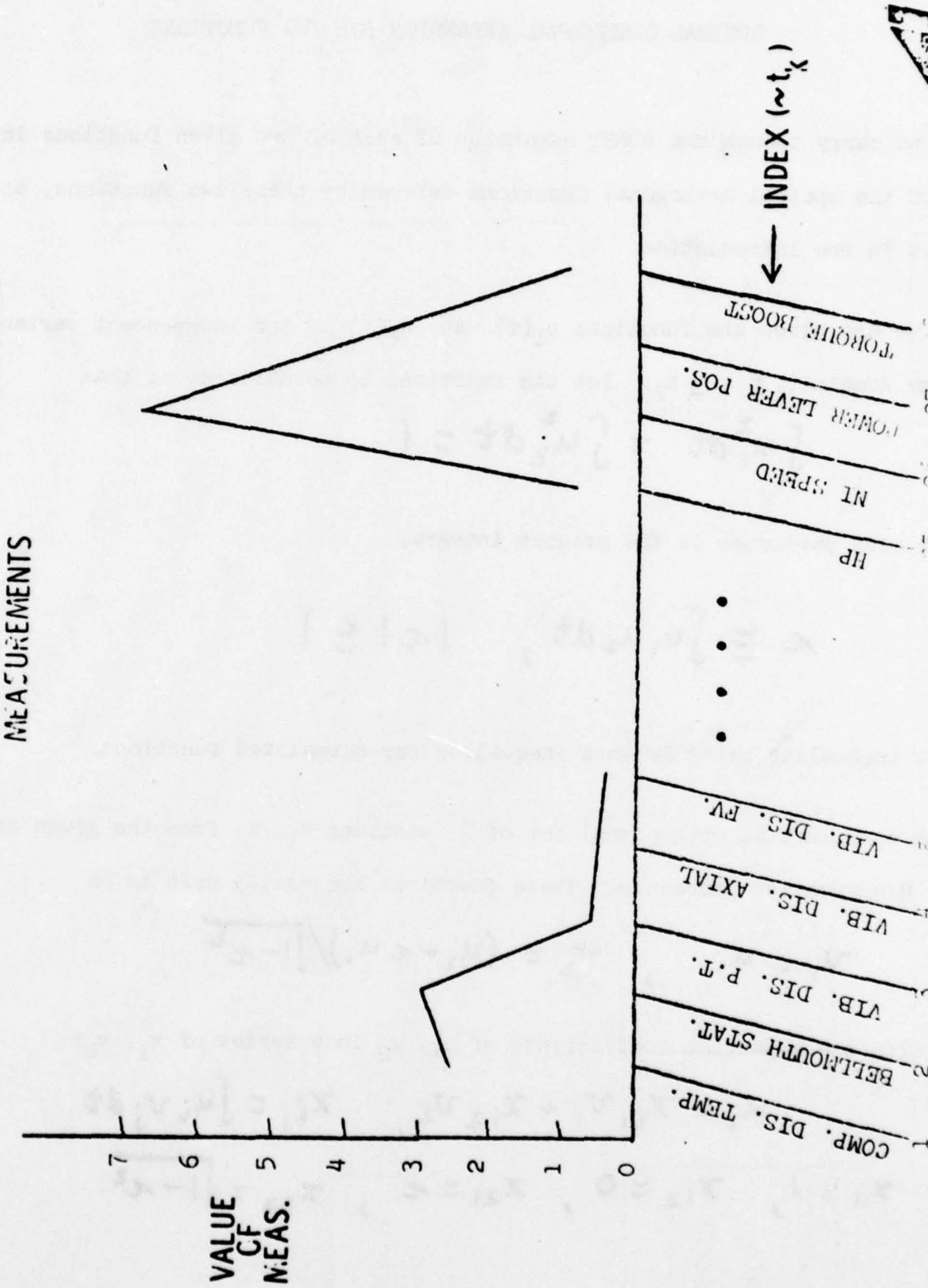
FIGURE 7

ADAPT - Techniques Programmed For Empirical Signature Analysis



1-1698

FIGURE 8
FORMULATION OF OBSERVABLE HISTORY FROM DISCRETE MEASUREMENTS



1-2555
6-69

APPENDIX B

OPTIMAL ORTHOGONAL EXPANSION FOR TWO FUNCTIONS

We wish to carry though the ADAPT expansion of each of two given functions in the series of the optimal orthogonal functions defined by these two functions, as described in the Introduction.

Suppose we are given the functions $u_1(t)$ and $u_2(t)$ of the independent variable t, over some domain $t_1 \leq t \leq t_2$. Let the functions be normalized, so that

$$\int u_1^2 dt = \int u_2^2 dt = 1$$

Then the only parameter is the product integral

$$c \equiv \int u_1 u_2 dt , \quad |c| \leq 1$$

the last inequality being Schwarz' inequality for normalized functions.

First we construct an orthonormal set of 2 functions v_1, v_2 from the given ones by the Gram-Schmidt procedure. These functions are easily seen to be

$$v_1 = u_1 \quad , \quad v_2 = (u_2 - c u_1)/\sqrt{1-c^2}$$

We now find the expansion coefficients of u_1, u_2 in a series of v_1, v_2:

$$u_i = x_{i1} v_1 + x_{i2} v_2 , \quad x_{ij} = \int u_i v_j dt$$

$$x_{11} = 1 , \quad x_{12} = 0 , \quad x_{21} = c , \quad x_{22} = \sqrt{1-c^2}$$

The optimal orthogonal functions are now obtained by finding the eigenvalues and eigenvectors $\underline{d}$ of the two-by-two matrix

$$S = \frac{1}{2}\left[x_{1i}x_{1j} + x_{2i}x_{2j}\right]$$

(the factor in front corresponds to weighing by dividing by the number of functions, in our case 2.) They are easily found to be

$$\lambda_1 = \frac{1}{2}(1+|c|) \quad , \quad \lambda_2 = \frac{1}{2}(1-|c|)$$

$$\underline{d}_1 = (\sqrt{\lambda_1}, \sqrt{\lambda_2}) \quad , \quad \underline{d}_2 = (\sqrt{\lambda_2}, -\sqrt{\lambda_1})$$

The eigenvectors are the expansion coefficients of the optimal orthogonal functions h_1, h_2 in a series in v_1, v_2, i.e.,

$$h_i = d_{i1}v_1 + d_{i2}v_2 \quad , \quad \underline{d}_i = (d_{i1}, d_{i2})$$

Returning to the original u functions we find the associated optimal functions to be

$$h_1 = \frac{1/2}{\sqrt{\lambda_1}}\left(u_1 + \frac{c}{|c|}u_2\right) \quad , \quad h_2 = \frac{1/2}{\sqrt{\lambda_2}}\left(u_1 - \frac{c}{|c|}u_2\right)$$

and the expansions of the u functions in them are

$$u_1 = \sqrt{\lambda_1}\, h_1 + \sqrt{\lambda_2}\, h_2 \quad , \quad u_2 = \frac{c}{|c|}\left(\sqrt{\lambda_1}\, h_1 - \sqrt{\lambda_2}\, h_2\right)$$

1-2555
6-69

It is sufficient to discuss the case of $\kappa \geq 0$ because if $\kappa < 0$, a change in the sign of u_2 returns to the first case. We note that the optimal function h_1 is proportional to the average of the input functions. The average is intuitively the best single function to represent two functions, so we see the best single function is associated with the larger eigenvalue λ_1. The optimal function associated with λ_2 is proportional to the difference of the given functions.

We also note that

$$\lambda_1 + \lambda_2 = 1, \quad \lambda_1 - \lambda_2 = \kappa = \int u_1 u_2 \, dt \leq 1$$

The decrease in the eigenvalue from the first to the second is the product integral of the two functions. If the functions are closely correlated one would expect κ to be near unity, and λ_2 would be much less than λ_1. But if the functions are nearly uncorrelated one would expect κ to be small, and there is only a slight decrease in the eigenvalue, going from the larger to the smaller. Thus the rate of decrease of eigenvalues can be associated with the degree of correlation of the input functions.